Metabolomics approach to growth-age discrimination in mountain cultivated ginseng (Panax ginseng C. A. Meyer) using UPLC-Q-TOF/MS

Gan Chen[a], Hong Zhang[a], Jiaming Jiang[a], Simin Chen[a], Hongmei Zhang[a], Gongmin Zhang[a], Changwu Zheng[a, *], and Hongxi Xub[, *]

[a] *School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, P. R. China*

[b] *Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, P. R. China*

**Correspondence author: Changwu Zheng, E-mail: zhengcw@shutcm.edu.cn;*

Hongxi Xu, E-mail: xuhongxi88@gmail.com

Abstract:

Introduction: Ginseng is a famous herbal medicine and has been a top-ranking subject for scientific research in both the medical and health care fields. However, quality control of ginseng is being challenged due to the difficulty of defining the age of ginseng and the serious adulteration phenomenon in the market. Mountain cultivated ginseng (*Panax ginseng* C. A. Meyer) (MCG) is usually harvested at the age of 10 years old and above. At an age of greater than 15 years old, it is typically termed wild ginseng. The medicinal and commercial value of MCG increases with age. The aim of this study is to differentiate MCG of different ages, because there is a serious issue with low-aged MCG being falsely represented as high-aged MCG in the market.

Materials and Methods: UPLC-Q-TOF/MS was used to analyze 98 MCG samples. A principal component analysis (PCA) and partial least squares discrimination analysis (PLS-DA) were used to find patterns between samples were used to select influential components. Application models were developed using machine learning to identify MCGs of different ages.

Results: The nontargeted metabolomic analysis using UPLC-Q-TOF/MS revealed that the MCG samples aged 4-20 years old were clearly divided into three groups. Twenty-two potential age-dependent biomarkers allowing for differentiation between the three sample groups were discovered. Three models of machine learning are then used to predict new samples, and the optimal prediction model is selected by parameters such as accuracy, recall, precision and F1 score.

Conclusions: After the MCG samples were separated into distinct age phases, some biomarkers could be used to determine the age phases. These biomarkers were analyzed thoroughly for variation trends, and machine learning models established based on the screened biomarkers was successfully used to predict the age groups of the new samples.